UNITED STATES PATENT APPLICATION

FOR

METHOD AND APPARATUS FOR UNSCHEDULED FLOW CONTROL IN PACKET FORM

INVENTORS:

Edmund G. Chen
Ravikrishna Cherukuri
Ruchi Wadhawan

Prepared by:

Blakely, Sokoloff, Taylor & Zafman
12400 Wilshire Boulevard
Seventh Floor
Los Angeles, California 90025
(408) 720-8598

Attorney's Docket No. 004906.P084

EXPRESS MAIL CERTIFICATE OF MAILING

"Express Mail" mailing label number: **EL863955377US** Date of Deposit: **December 7, 2001**

I hereby certify that I am causing this paper or fee to be deposited with the United States Postal Service "Express Mail Post Office to Addressee" service on the date indicated above and that this paper or fee has been addressed to the Commissioner for Patents, Washington, D. C. 20231

Deborah Roszkowski
(Typed or printed name of person mailing paper or fee)

Deborah Roszkowski
(Signature of person mailing paper or fee)

12-7-01
(Date signed)

METHOD AND APPARATUS FOR UNSCHEDULED FLOW CONTROL IN PACKET FORM

FIELD OF THE INVENTION

[0001] The invention relates to the field of networking. More specifically, the invention relates to the transmission of flow control between two integrated circuit chips.

BACKGROUND OF THE INVENTION

[0002] Networks are comprised of various network elements, which process and transmit traffic (i.e., network packets such as Internet Protocol ("IP") packets, Ethernet packets, etc.) flowing through the network. Each network element may comprise a number of line cards (which include one or more ports and chips) and control cards (which include one or more chips) to transmit the traffic through the network element. Communications within a network element requires the various chips within a network element to communicate information.

[0003] Flow control is a technique for ensuring that a transmitting chip (e.g., a chip on a line card) does not overwhelm a receiving chip with data. Flow control operates to slow or temporarily halt transmission from the source, allowing the recipient, for example, another chip (e.g., another chip on the line card) to unload any stored data. Specifically, a receiving chip can transmit a flow control signal to a transmitting chip, and the transmitting chip can adjust its rate of data transmission accordingly. One such flow control scheme used for chips processing IP packet traffic is the Optical Internetworking Forum System Packet Interface Level 4 Phase 2: OC-192 System Interface for Physical and Link Layer Devices (January 2001) (the "OIF SPI-4.2.2 standard"). This standard is TDM-based (it combines different data streams, such that each data stream is assigned a time slot within a TDM data stream). The time slots are referred to as operating according to a calendar.

[0004] Although the OIF SPI-4.2 standard has a low pin count, it suffers from several disadvantages. For instance, network packet transmissions are typically sent in bursts. Because the OIF SPI-4.2 standard does not permit flow control to be transmitted when needed (instead, it requires adherence to the TDM-based calendar), it does not model the bursty nature of IP packet traffic well and therefore, requires additional buffering. In addition, the bandwidth of network packet transmissions can vary widely on a per channel basis. However, since the OIF SPI-4.2 standard uses fixed time slots, variations of bandwidth allocation are difficult to program (one or more time slots must be selected and programmed). In addition, the OIF SPI-4.2 standard requires the programming (initially as well as to add/delete channels) of both the sender and receiver. Further, for the addition or deletion of ports, the OIF SPI-4.2 standard requires that the programming of the calendar of the sender and recipient be in sync. Also, accumulated state information must be maintained for each packet buffer to allow for the delay between when a network packet is received and when flow control information can be transmitted.

SUMMARY OF THE INVENTION

[0005] A method and apparatus for transmitting unscheduled flow control, in packet form, between two chips are described. In one embodiment, a method includes reading a status of a buffer used to receive network packets transmitted from a different chip. The method further includes transmitting to said different chip an unscheduled flow control packet including information about the status of the buffer.

[0006] In an embodiment, a chip includes a packet buffer to store network packets transmitted from a different chip, wherein the packet buffer is associated with one or more of a plurality of ports through which the network packets travel. The chip also includes control circuitry, coupled with a packet data bus to receive said network packets from the different chip, and coupled with an unscheduled flow control packet bus to generate and transmit unscheduled flow control packets to the different chip,

wherein the unscheduled flow control packets contain information relating to the packet buffer.

BRIEF DESCRIPTION OF THE DRAWINGS

[0007] Embodiments of the invention may be best understood by referring to the following description and accompanying drawings that illustrate such embodiments. The numbering scheme for the Figures included herein are such that the leading number for a given element in a Figure is associated with the number of the Figure. For example, packet sender chip 102 can be located in Figure 1. However, element numbers are the same for those elements that are the same across different Figures.

[0008] In the drawings:

[0009] **Figure 1** is a block diagram illustrating portions of a packet sender chip coupled with a packet recipient chip, according to embodiments of the invention;

[0010] **Figure 2A** illustrates a flow diagram for the transmission of unscheduled flow control packets, according to embodiments of the invention;

[0011] **Figure 2B** illustrates an unscheduled flow control packet according to embodiments of the invention; and

[0012] **Figure 3** illustrates a flow diagram for the receipt of unscheduled flow control packets and the modification of a rate of transmission of network packets, according to embodiments of the invention.

DETAILED DESCRIPTION

[0013] A method and apparatus for transmitting unscheduled flow control, in packet form, between two chips are described. In the following description, for purposes of

explanation, numerous specific details are set forth in order to provide a thorough understanding of the invention. It will be evident, however, to one skilled in the art that the present invention may be practiced without these specific details. In other instances, well-known circuits, structures, and techniques have not been shown in detail in order not to obscure the invention.

[0014] **Figure 1** is a block diagram illustrating portions of a packet sender chip coupled with a packet recipient chip, according to embodiments of the invention. In particular, Figure 1 illustrates packet sender chip 102 and packet recipient chip 104, which are coupled together via packet data bus 106 and unscheduled flow control bus 108. Together, packet sender chip 102 and packet recipient chip 104 are, respectively, any sender and recipient of packets in a network element for which flow control is required. For example, packet sender chip 102 and packet recipient chip 104 could be a packet processor, a chip that interfaces the packet processor on a given line card to other line cards, a chip to interface the packet processor to the physical ports, etc.

[0015] Packet sender chip 102 can transmit network packets to packet recipient chip 104 over packet data bus 106. While in one embodiment packet data bus 106 is a serial bus, in alternative embodiments packet data bus 106 may be any other kind of bus for the transmission of network packets (e.g., a combination of busses, which when grouped together, provide for the transmission of network packets). In an embodiment, the network packets transmitted over packet data bus 106 can include Internet Protocol ("IP"), asynchronous transfer mode ("ATM"), frame relay, voice-over IP ("VOIP"), point-to-point protocol ("PPP") packets, etc. Packet recipient chip 104 can transmit unscheduled flow control packets to packet sender chip 102 over unscheduled flow control bus 108. While in one embodiment unscheduled flow control bus 108 is a serial bus, in alternative embodiments unscheduled flow control bus 108 may be any other kind of bus for the transmission of network packets (e.g., a combination of busses,

which when grouped together, provide for the transmission of unscheduled flow control packets).

[0016] Packet recipient chip 104 comprises packet buffers 110A-N (collectively, "packet buffer 110") to temporarily store the network packets received from packet sender chip 102. The network packets are each associated with a port through which the network packets are transmitted out of the network element. The ports may be physically further downstream or coupled with packet recipient chip 104. In one embodiment, packet buffer 110 can operate in two modes: 1) in the port specific mode there are dedicated packet buffers for each particular port through which network packets will flow; and 2) in the aggregated port mode there are shared packet buffers for all of the ports. However, alternative embodiments of the invention can support more, less, and/or different buffer modes.

[0017] Packet recipient chip 104 further comprises control circuitry 116, which is coupled with unscheduled flow control bus 108 and packet data bus 106, to receive the network packets from packet sender chip 102 and to generate and transmit unscheduled flow control packets to packet sender chip 102. Packet sender chip 102 further comprises flow control logic 112, coupled with unscheduled flow control bus 108, to receive the unscheduled flow control packets from packet recipient chip 104. Packet sender chip 102 further comprises network packet logic 114, coupled with packet data bus 106 and flow control logic 112, to modify, in response to the unscheduled flow control packets, a rate at which the network packets are transmitted to packet recipient chip 104.

[0018] In an embodiment, packet recipient chip 104 further transmits the network packets received from packet sender chip 102 to another chip (e.g., another chip on the line card). In such an embodiment, flow control information may be required between packet recipient chip 104 and the chip to which it transmits network packets. In such

an embodiment, packet recipient chip 104 is coupled with the other chip via another unscheduled flow control bus and another packet data bus, which operate in a similar manner as unscheduled flow control bus 108 and packet data bus 106, respectively. In such an embodiment, packet recipient chip 104 further comprises flow control logic and network packet logic, which operate in a similar manner as flow control logic 112 and network packet logic 114, respectively. In such an embodiment, the other chip comprises packet buffers and control circuitry, which operate in a similar manner as packet buffers 110A-N and control circuitry 116, respectively.

[0019] **Figure 2A** illustrates a flow diagram for the transmission of unscheduled flow control packets, according to embodiments of the invention. In one embodiment, the flow diagram 201 is performed each time packet buffer status information needs to be sent (i.e., it need not follow a calendar). Checking whether status information for a given packet buffer (or the aggregate packet buffer) needs to be sent can be done at any number of different times (e.g., each time a network packet is received in packet buffer, when the packet buffer is idle for a predetermined period of time, etc.).

[0020] Flow diagram 201 begins at process block 202, where the status of the packet buffer is determined. At process block 204, an unscheduled flow control packet is allocated. **Figure 2B** illustrates an unscheduled flow control packet 220 according to embodiments of the invention. In one embodiment, unscheduled flow control packet 220 comprises command field 222 and port identification field 224. Command field 222 comprises the command to packet sender chip 102, e.g., idle, full, reserved, Xon, Xoff, credit acknowledgements, etc. In one embodiment, the command field 222 supports OIF SPI-4.2 commands. Table 1, appearing below, includes supported OIF SPI-4.2 commands.

Field	Bit	Description
Type	3:2	Command Type 2'b00: Idle 2'b01: Reserved 2'b10: Max 1 Ack 2'b11: Max 2 Ack
DIP-2	1:0	DIP-2 Code
Port In	n:0	Port Identification

Table 1 : Flow Control Commands

Port identification field 224 comprises the address of the port through which network packets will travel or an aggregated packet buffer identifier (as discussed below).

[0021] Returning to Figure 2A, at process block 206, the command field is filled with the command for network packet transmissions intended for the packet buffer. At decision block 208, it is determined whether the packet buffer is in port-specific mode or aggregated port mode. If the packet buffer is in port-specific mode, the port identification field in the unscheduled flow control packet is filled with the address of the port to which the packet buffer is associated, at process block 210. If the packet buffer is in aggregated port mode, the port identification field in the unscheduled flow control packet is filled with an aggregated packet buffer identifier, at process block 212. In an embodiment, the aggregated packet buffer identifier is a predetermined combination of bits that indicates that packet buffer 110 is being used in the aggregate for all ports. Unlike the prior art, where specific time slots are allocated for a particular port, the unscheduled flow control packets may be sent (if port-specific mode, then in regard to any port as well) at any time.

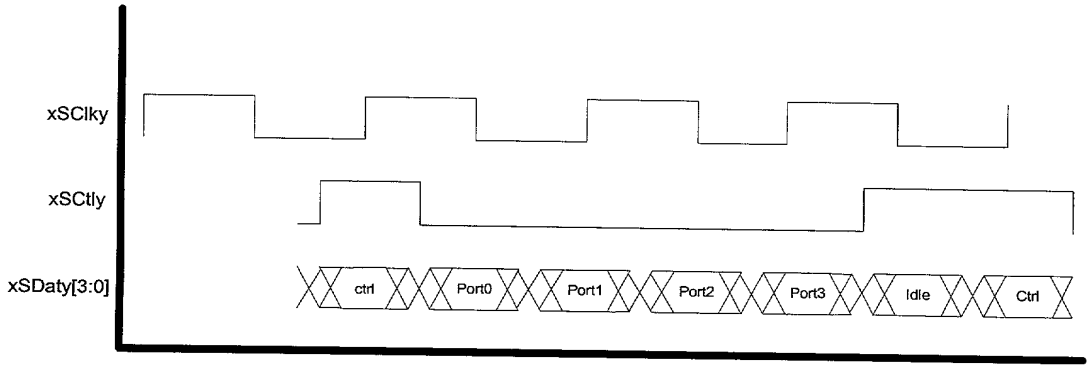
[0022] At process block 214, the unscheduled flow control packet is transmitted to packet sender chip over unscheduled flow control bus. While in one embodiment, the

unscheduled flow control packets may be variable in size (depending on the contents of the command and port identification fields), alternative embodiments use a fixed size. In an embodiment, unscheduled flow control bus 108 is comprised of five lines: four lines for port information and one line for control information. However, other embodiments may include a different number of lines and a different allocation among the lines. In addition, one embodiment supports both unscheduled flow control packets and the OIF SPI-4.2 standard. This embodiment includes a mechanism to indicate which mode to operate in: the unscheduled flow control packet mode or the OIF SPI-4.2 standard mode. In addition, this embodiment uses a subset of the pins used for the unscheduled flow control packet mode for the OIF SPI-4.2 standard mode. Table 2, appearing below, includes pins used for unscheduled flow control packet mode.

Pin Name	Number Pins	I/O Type	Description
Clk	2	LVDS	Status Clock
Ctl	2	LVDS	Status Control
Dat[3:0]	8	LVDS	Status Data
TOTAL	12		

Table 2 : Enhanced Flow Control Pins

[0023] Timing Chart 1 illustrates the flow control timing of the pins included in Table 2.



Timing Chart 1: Flow Control Timing

Specifically, the control line and two of the data lines are used when operating in the OIF SPI-4.2 standard mode.

[0024] In one embodiment, unscheduled flow control bus 108 permits unscheduled flow control packets to be transmitted to packet sender unit 102 faster than network packets are transmitted to packet recipient unit 104, thus negating the need for storing any accumulated packet buffer state. In an embodiment, unscheduled flow control bus 108 is a serial bus, but embodiments of the present invention are not so limited. For example, in another embodiment, unscheduled flow control bus 108 is a multi-bit parallel bus. In an embodiment, unscheduled flow control packets take five clock cycles for transmission, and as stated previously, are transmitted whenever flow control information is necessary.

[0025] Transmitting flow control information in a packet form (that identifies the packet buffer) whenever flow control is necessary (i.e., unscheduled) provides several distinct advantages. For instance, such unscheduled flow control packets model well

the bursty nature of network packet transmissions. In addition, the allocation of bandwidth between channels is achieved merely by adjusting the frequency of flow control packet transmissions for different packet buffers on an as needed basis. Further, a calendar on both the flow control sender and receiver need not be programmed and reprogrammed. Rather the flow control sender generates flow control packets (that identify the packet buffer to which they apply) on an as needed basis (synchronication is not an issue).

[0026] **Figure 3** illustrates a flow diagram for the receipt of unscheduled flow control packets and the modification of a rate of transmission of network packets, according to embodiments of the invention. In particular, Figure 3 illustrates flow diagram 301, which is executed each time an unscheduled flow control packet is received by the packet sender chip. At process block 302, the unscheduled flow control packet is received by the packet sender unit. At process block 304, it is determined if the packet buffer is being operated in port-specific mode. In one embodiment, the packet sender chip determines this by examining the port identification field (which was filled accordingly, in flow diagram 201). Specifically, the port field will contain either the address of a specific port (port-specific mode) or the aggregate packet buffer identifier. If operating in port-specific mode, the packet sender chip modifies the rate at which network packets associated with the particular port referenced in the unscheduled flow control packet are sent (process block 306). If operating in the aggregated port mode, then the packet sender chip modifies the rate at which all network packets are transmitted to the packet recipient chip regardless of their port association (process block 308).

[0027] The line cards and control cards included in the different network elements include memories, processors, and/or Application Specific Integrated Circuits ("ASICs"). Such memory includes a machine-readable medium on which is stored a set of instructions (i.e., software) embodying any one, or all, of the methodologies

described herein. Software can reside, completely or at least partially, within this memory and/or within the processor and/or ASICs. For the purposes of this specification, the term "machine-readable medium" shall be taken to include any mechanism that provides (i.e., stores and/or transmits) information in a form readable by a machine (e.g., a computer). For example, a machine-readable medium includes read only memory (ROM), random access memory (RAM), magnetic disk storage media; optical storage media, flash memory devices, electrical, optical, acoustical, or other form of propagated signals (e.g., carrier waves, infrared signals, digital signals, etc.), etc.

[0028] Thus, a method and apparatus for transmitting unscheduled flow control, in packet form, between two chips have been described. Although the present invention has been described with reference to specific exemplary embodiments, it will be evident that various modifications and changes may be made to these embodiments without departing from the broader spirit and scope of the invention. For example, different input formats may be used. Accordingly, the specification and drawings are to be regarded in an illustrative rather than a restrictive sense.